Компьютерное зрение (Computer Vision), а также иногда машинное зрение (Machine Vision) – научная область, занимающаяся исследованиями в области автоматической фиксации и разного рода обработки изображений (обнаружение, отслеживание, идентификация) с помощью компьютера.

Одни из первых попыток создать такие системы были предприняты в 1960-х годах, но из-за низкой мощности компьютеров того времени эта область долго не была достаточно исследованной. Современные скоростные процессоры, дешевая память на диске и высококачественные камеры и высокая пропускная способность каналов позволяют добиться значительных результатов в этой области, особенно в связи с последними исследованиями в области машинного обучения и, в частности, нейронных сетей.

Рост качества технологий компьютерного зрения позволил активно их использовать в разных областях промышленности и бизнеса. К примеру, многие современные конвейеры оснащены автоматическими механизмами проверки качества деталей, корректной маркировки, выравнивания деталей и других задач. Примером такого подхода может служить индустрия микрочипов, где камеры автоматически контролируют размещение кремниевых пластин и планировку интегральных схем. Можно перечислить большое количество областей цифровой экономики, в которых применяется современное компьютерное зрение:

- Системы “умный город”.

- Инфраструктурные и транспортные системы с контролем и балансировкой нагрузки.
- Автомобильные системы, в том числе беспилотные автомобили.
- Беспилотные летательные аппараты, в том числе дроны.
- Фармацевтика и медицина.
- Видеоаналитики.
- Распознавание человека и предметов.
- Идентификация лиц.
- Оценка сцены, времени, места.

Основная цель компьютерного зрения – получение полезной информации из изображения или серии изображений. Задачами могут быть такие, как:

- калибровка оптических систем, настройка и синхронизация настроек камер,
- определение движения объектов (Motion Tracking),
- задачи распознавания объектов,
- задачи реконструкции сцены (обычно 3D-сцены из одного или нескольких 2D-кадров),
- задачи сравнения изображений и идентификации изменений.

За последние десять лет компьютерное зрение в пищевой промышленности пережило заметный рост исследований. Согласно данным, полученным из WoS, с 2012 по 2022 гг. было опубликовано около 222 статей (123 исследовательских и 61 обзорная статья, а также 38 сборников/глав книг/редакционных материалов). Такой

рост публикаций свидетельствует о возросшем интересе и исследовательской активности к использованию компьютерного зрения в пищевой промышленности.

Компьютерное зрение может быть использовано для персонализации пищевых продуктов под отдельного потребителя. Например, он может легко определить диету
Технология искусственного интеллекта, облегчающая мониторинг в режиме реального времени, стала важным инструментом в области автоматизации пищевой промышленности и сельского хозяйства.
Методы на основе СНС Сильные стороны СНС и модели трансформатора объединены в семействе подходов к глубокому обучению, известных как методы на основе трансформаторов СНС, которые обеспечивают современную производительность в различных приложениях компьютерного зрения (Moutik et al., 2023). Подход, основанный на СНС-трансформаторе, использует СНС для извлечения визуальной информации из входных изображений и трансформаторы для моделирования взаимосвязей между этими функциями, сочетая возможности обеих технологий (рис. 5C). Различные задачи компьютерного зрения, включая классификацию изображений, обнаружение объектов и семантическую сегментацию, были успешно решены с помощью этих методов (Carion et al., 2020), как показано на рис.

CNN-transformer-based methods CNN and transformer model strengths are combined in a family of deep learning approaches known as CNN-transformer-based methods, which enable state-of-the-art performance in a variety of computer vision applications (Moutik et al., 2023). The CNN-transformer-based approach uses CNNs to extract visual information from input images and transformers to model the relationships between these features, combining the capabilities of both technologies (Fig. 5C). Various computer vision tasks, including image classification, object detection, and semantic segmentation, have been successfully accomplished using these techniques (Carion et al., 2020) as shown in Fig. 6.

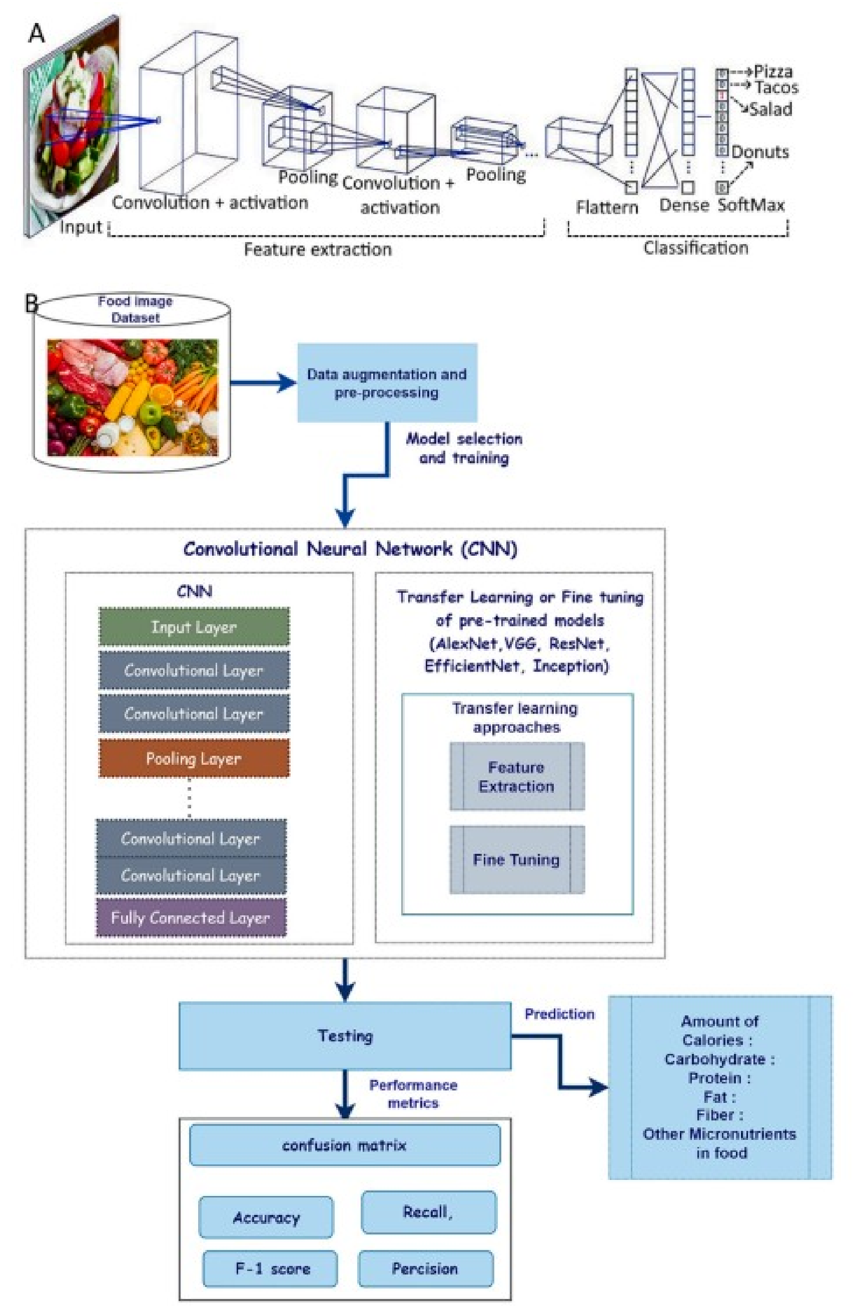


Fig. 4. Framework of CNN-based model. (A) automated dietary assessment system and (B) food image classification and nutrient identification process. Panels A and B are adapted with permission (copyright © 2020 by editors and copyright © 2023 by authors) from Kiourt, Pavlidis, and Markantonatou (2020) and Kaur, Kumar, and Gupta (2023), respectively.

[1] -- Sushant Kaushal, Dushyanth Kumar Tammineni, Priya Rana, Minaxi Sharma, Kandi Sridhar, Ho-Hsien Chen Computer vision and deep learning-based approaches for detection of food nutrients/nutrition: New insights and advances [Текст] / Sushant Kaushal, Dushyanth Kumar Tammineni, Priya Rana, Minaxi Sharma, Kandi Sridhar, Ho-Hsien Chen // Trends in Food Science & Technology. — 2024. — № 146. —

Компьютерное зрение – часть области искусственного интеллекта (AI) Как и для других задач, для компьютерного зрения есть своя версия теста

Тьюринга, т.е. компьютер должен ответить на любой вопрос про изображение, на который может ответить человек. Человек и компьютер могут отвечать на следующие вопросы про изображение: 1) Что и где находится на изображении? – задача детекции 2) Вопрос о свойстве объектов и их атрибутов – задача классификации в зависимости от ответа 3) Какой формы и какого размера объект? – задача метрического зрения, сложной версией которой является построение 3-D модели объекта

-- Ускорение задач мультимедийной геолокации для судебного преследования за торговлю людьми. для эффективной борьбы с торговлей людьми, сексуальной эксплуатацией детей и другими противоправными действиями.

-- transfer learning, image inhancement, Feature Matching and Tracking, image Segmentation, Deep Learning and Neural Network - Computer Vision Techniques Related to Multimedia Processing -- Object Detection – yolo, cnn, r-cnn

--Мультимедийная геолокация — это процесс оценки реального местоположения, в котором сделан мультимедийный файл, например, изображение или видео [3]. Аналогичным образом, геолокация изображений требует умения извлекать визуальные атрибуты из изображений и понимания геопространственного распределения этих атрибутов [45].

Zhang et al. [70] исследовали, способствует ли пространственная корреляция между визуальным и текстовым контентом ориентиров развитию мультимедиа. Они начинают с внедрения сети уточнения признаков, которая изучает различающие и надежные визуальные представления, комбинируя глобальные и локальные характеристики. Во-вторых, чтобы повысить производительность поиска, они используют архитектуру обучения мультимодального классификатора, которая включает визуальные и текстовые входные данные. Их эксперименты на реальных наборах данных демонстрируют превосходство предложенного подхода по сравнению с существующими методами Аналогичным образом, Nilwong et al. [41] предложили два метода локализации на открытом воздухе, основанные на глубоком обучении и идентификации ориентиров. Первый метод основан на обнаружении ориентиров на собранном изображении с помощью более быстрой регионально-сверточной нейронной сети (Faster R-CNN). Затем, используя наблюдаемые ориентиры, нейронная сеть прямого распространения (FFNN) обучается для оценки координат местоположения и ориентации по компасу, как показано на рисунке 5. Во втором методе используется архитектура CNN, как показано на рисунке 6. Оба результата состоят из углов широты и долготы, а также ориентации магнитного компаса. Метод, основанный на Faster R-CNN, превзошел метод на основе CNN при оценке широты и долготы, но пострадал от снижения точности компаса

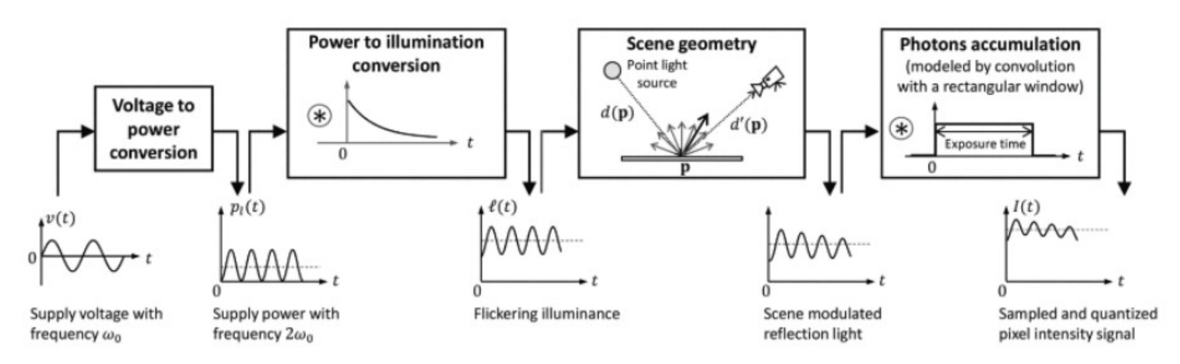



Figure 7: An illustration of the ENF embedding process [69]
Рисунок 7: Иллюстрация процесса встраивания ENF [69]

Снимки сверху легко доступны и обеспечивают плотное покрытие при все более высоком разрешении. Однако мелкозернистые признаки иногда трудно определить сверху вниз. Фотографии с геотегами на уровне земли разбросаны редко, но содержат семантически богатую информацию с высоким разрешением. Комбинирование и сопоставление фотографий с высоты птичьего полета и с уровня улицы с использованием локальных характеристик, таких как выходные данные SIFT, сложно из-за широко варьирующейся точки обзора, масштаба, освещенности, модальности датчика и времени съемки, но с помощью передовых рабочих процессов с использованием компьютерного зрения это осуществимо. При наличии дополнительных снимков с уровня земли сегментация изображений сверху может привести к значительному повышению точности [56]. Lefèvre et al. [82] обратили внимание на сложность объединения и интерпретации спутниковых снимков и снимков улиц для получения целостного понимания сцен. Авторы разработали фреймворк, который сочетает в себе методы глубокого обучения и алгоритмы компьютерного зрения для анализа сцен и извлечения объектов со спутника на уровень улицы для установления соответствий между различными видами. Это позволяет интегрировать информацию из обоих источников для более точной оценки местоположения. Аналогичным образом, Li et al. [81] подходят к этой проблеме в три этапа; 1) построить модель глубокого обучения для изучения семантических представлений изображений, .2) извлечь семантические признаки из эталонного и целевого изображений с помощью обученной модели глубокого обучения, и .3) предложить систему сопоставления шаблонов, которая исследует семантические свойства, чтобы выявить соответствия между эталонными и целевыми изображениями, что позволит точно регистрировать. Это обеспечивает более быстрый анализ изображений с помощью дистанционного зондирования и обработку конвейеров, а также открывает новые пути регистрации на основе обучения.

Overhead imagery is readily accessible and offers dense coverage at progressively high resolutions. However, fine-grained features are sometimes difficult to define from the top down. Geotagged ground-level photos are sparsely scattered yet capture high-resolution, semantically rich information. Combining and matching overhead and streetlevel photographs is difficult using local features such as the output of the SIFT due to the widely variable viewpoint, scale, illumination, sensor modality, and acquisition time, but with advanced computer vision-aided workflows, it is feasible. When supplemental ground-level photos are available, overhead image segmentation can result in

considerable accuracy increases [56]. Lefèvre et al. [82] addressed the difficulty of combining and interpreting satellite images and street-level imagery to get a holistic understanding of scenes. The authors developed a framework that combines deep learning techniques and computer vision algorithms to analyze scenes and extract features from satellite to street level to establish correspondences between different views. This allows the integration of information from both sources for more accurate location estimation. Similarly, Li et al. [81] approach this problem in three steps; 1) construct a deep learning model to learn semantic representations of images, .2) extract semantic features from the reference and target images using the trained deep learning model, and .3) propose a template matching system that examines semantic properties in order to uncover correspondences between reference and target images, allowing for accurate registration. This provides faster remote sensing picture analysis and pipeline processing, while also allowing for new paths in learning-based registration.

Чой Джэён [7] полагался на наличие видимых структур с обоих ракурсов, чтобы выровнять две плоскости таким образом, чтобы свести к минимуму ошибку выравнивания. Изменяя поле зрения (FOV) на совмещенном спутниковом снимке, авторы извлекали поле зрения из изображения, чтобы воспроизвести вид с уровня земли в определенном направлении обзора. Затем определяется горизонт изображения и выбирается область нулевой плоскости для сопоставления. Workman et al. [56] использовали геометрический механизм внимания, который учитывает географическую связь между пикселями на снимке на уровне земли и географическим положением для сегментации в дистанционном зондировании. Это можно рассматривать как перекрестную геолокацию, когда оценка геолокации изображения на уровне земли становится возможной благодаря привязке к снимкам сверху. Кроме того, 13 стандартных трекеров, используемых в традиционном визуальном отслеживании, были тщательно оценены Shao et al. [84]. Согласно их экспериментальным выводам, большинство передовых алгоритмов слежения полагаются в первую очередь на яркость, цвет или свёрточные характеристики, и поэтому не могут следовать за объектами спутникового видео из-за недостаточности этих атрибутов представления. Для решения этой проблемы авторы предлагают использовать скоростной корреляционный фильтр путем построения специального фильтра корреляции ядра для слежения за объектами спутникового видео, используя как скоростной признак, так и инерционный механизм

Jaeyoung Choi [7] relied on the availability of visible structures from both views to align the two planes in such a way that the alignment error is minimized. By changing the field of view (FOV) on the co-located satellite picture, the authors retrieved the FOV from the image to replicate a ground-level view in a certain viewing direction. The horizon of the image is then identified, and a ground plane region is chosen for matching. Workman et al. [56] used a geometry-aware attention mechanism that takes the geographical link between pixels in a ground-level picture and a geographic location into account for segmentation in remote sensing. This can be seen as cross-view geolocation, where ground-level image geolocation estimation is made possible through overhead imagery reference. In addition to this, 13 standard trackers used in traditional visual tracking were thoroughly evaluated by Shao et al. [84]. According to their experimental findings, the majority of cutting-edge tracking algorithms rely primarily on brightness, color, or convolutional features, and are therefore unable to follow satellite video objects because these representation attributes

are insufficient. The authors suggest a velocity correlation filter approach to get over this problem by building a specific kernel correlation filter for the satellite video target tracking using both a velocity feature and an inertia mechanism.

7. Jaeyoung Choi GFe. Multimodal Location Estimation of Videos and Images. 1 ed.; Springer International Publishing; 2015. ISBN 978-3- 319-09860-9,978-3-319-09861-6. URL: http://gen.lib.rus.ec/book/index.php?md5=5c5f9e64ebb1bda87b88dfe1efdd22f5. doi:10.1007/978-3-319-09861-6.

84. Shao J, Du B, Wu C, Zhang L. Tracking objects from satellite videos: A velocity feature based correlation filter. IEEE Transactions on Geoscience and Remote Sensing 2019;57(10):7860–7871. doi:10.1109/TGRS.2019.2916953.

56. Workman S, Rafique MU, Blanton H, Jacobs N. Revisiting near/remote sensing with geospatial attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022:1778–1787.

81. Li L, Han L, Ding M, Cao H, Hu H. A deep learning semantic template matching framework for remote sensing image registration. ISPRS Journal of Photogrammetry and Remote Sensing 2021;181:205–217. URL: https://www.sciencedirect.com/science/article/ pii/S0924271621002446. doi:https://doi.org/10.1016/j.isprsjprs.2021.09.012.

82. Lefèvre S, Tuia D, Wegner JD, Produit T, Nassar AS. Toward seamless multiview scene analysis from satellite to street level. Proceedings of the IEEE 2017;105(10):1884–1899. doi:10.1109/JPROC.2017.2684300.

70. Zhang X, Wang S, Li Z, Ma S. Landmark image retrieval by jointing feature refinement and multimodal classifier learning. IEEE Transactions on Cybernetics 2018;48(6):1682–1695. doi:10.1109/TCYB.2017.2712798

45. Salem T, Workman S, Jacobs N. Learning a dynamic map of visual appearance. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020:12435–12444.

3. Glistrup M, Rudinac S, Jónsson B. Urban image geo-localization using open data on public spaces. 2022:50–56. URL: https://www.researchgate.net/publication/364252863_Urban_Image_Geo-Localization_Using_Open_Data_on_ Public_Spaces. doi:10.1145/3549555.3549589.

41. Nilwong S, Hossain D, Kaneko Si, Capi G. Deep learning-based landmark detection for mobile robot outdoor localization. Machines 2019;7(2). URL: https://www.mdpi.com/2075-1702/7/2/25. doi:10.3390/machines7020025.

--------------- Opeyemi Bamigbade , Mark Scanlon , John Sheppard Computer Vision for Multimedia Geolocation in Human Trafficking Investigation: A Systematic Literature Review [Текст] / Opeyemi Bamigbade , Mark Scanlon , John Sheppard // . — 2024.

# СИСТЕМЫ КОМПЬЮТЕРНОГО ЗРЕНИЯ: СОВРЕМЕННЫЕ ЗАДАЧИ И

МЕТОДЫ
АЛЕКСЕЙ ПОТАПОВ
#1 (49), 2014 CONTROL ENGINEERING РОССИЯ стр 21-26

Таким образом, в плане практического применения системы компьютерного зрения прошли ряд этапов: этап индивидуального решения (как в части аппаратного обеспечения, так и алгоритмов) конкретных задач; этап применения в профессиональных областях (в особенности в промышленности и оборонной сфере) с использованием спецпроцессоров, специализированные системы формирования изображений и алгоритмы, предназначенные для работы в условиях низкой априорной не определенности, однако эти решения допускали масштабирование;

Как видно, система машинного зрения включает следующие основные компоненты: • подсистему формирования изображений (которая сама может включать разные компоненты, например объектив и ПЗС- или КМОП-матрицу); • вычислитель; • алгоритмы анализа изображений, которые могут реализовываться программно на процессорах общего назначения, аппаратно в структуре вычислителя и даже аппаратно в рамках подсистемы формирования изображений.

Таким образом, современная робототехника требует решения широкого круга задач компьютерного зрения, включающего, в частности: • набор задач, связанных с ориентацией во внешнем пространстве (например, задачу одновременной локализации и картографирования — Simultaneous Localization and Mapping, SLAM), определением расстояний до объектов и т. д; • задачи по распознаванию различных объектов и интерпретации сцен в целом; • задачи по обнаружению людей, распознаванию их лиц и анализу эмоций

JIAQI LI , QI MIAO, ZHENG ZOU, HUAGUO GAO, LIXIAO ZHANG, ZHAOBO LI, AND NAN WANG A Review of Computer Vision-Based Monitoring Approaches for Construction Workers’ Work-Related Behaviors [Текст] / JIAQI LI , QI MIAO, ZHENG ZOU, HUAGUO GAO, LIXIAO ZHANG, ZHAOBO LI, AND NAN WANG // IEEE Access. — 2024. — № 12. — С. 7134-7155.
Обзор подходов к мониторингу рабочего поведения строителей на основе компьютерного зрения

Cai et al. [84] использовали Faster R-CNN для определения ориентации головы и тела строителей, а затем применили многозадачную обучающую сеть для оценки направления зрительного внимания рабочих. Yan et al. [39] сначала обнаружили рабочих с помощью Faster R-CNN, а затем применили 3D-оценку позы для определения их местоположения в трехмерном пространстве. Фундаментальная основа для внедрения отслеживания работников заключается в локализации и выявлении строительного персонала, что способствует более глубокому пониманию распределения рабочей силы и мобильности на строительных площадках. Алгоритм фильтрации Калмана нашел широкое применение в области слежения за объектами, о чем свидетельствует рассмотренная литература [109], [113], [120], [121], [122]. Фильтрацию Калмана принято комбинировать с алгоритмами обнаружения объектов. Например, Neuhausen et al. [120], [121] использовали YOLO v3 для обнаружения строителей, а затем

отслеживали их с помощью фильтра Калмана, демонстрируя устойчивость фильтра Калмана в компенсации ограничений YOLO v3. Yongyue et al. [122] предложили другой подход, объединив фильтрацию Калмана с алгоритмом оценки точек соединения отдельных поз Openpose, чтобы уменьшить рабочую нагрузку на ручное аннотирование. Аналогичных результатов можно достичь, комбинируя алгоритмы обнаружения объектов с другими методами. Son и Kim [63] предложили комбинацию YOLO v4 и Siamese Network для обнаружения и отслеживания строительных рабочих, достигнув точности 0,975. Wan et al. [106] усовершенствовали YOLO v5 с помощью механизма внимания, астропространственного пулинга и универсального апсамплинга для отслеживания строителей путем классификации касок, что, в свою очередь, выявляло несанкционированные вторжения. Анга и Чен [123] интегрировали Mask R-CNN, Matching и Rematching, что позволило отслеживать нескольких строительных рабочих на экране. Liu et al. [97] ввели многодоменное представление для CNN для улучшения эффектов слежения в сложных динамических сценах. Кроме того, исследователи расширили отслеживание в 3D-пространстве на основе 2D-отслеживания. Цай и Цай [81] предложили гибридный подход визуального слежения и радиолокации для решения проблемы легкой потери целей при визуальном распознавании. Ли и Парк [124] использовали две камеры для раздельного отслеживания и локализации, а затем Entity Matching для 3D-отслеживания. Сети LSTM могут анализировать входные данные с компонентом временных рядов. Для прогнозирования траекторий действий строителей Cai et al. [83] предложили контекстно-зависимый метод, основанный на LSTM. Аналогичным образом, Tang et al. [73] объединили LSTM с сетью плотности смеси для достижения прогнозирования траекторий действий строителей. Методы обнаружения и отслеживания на основе компьютерного зрения могут эффективно помочь руководителям проектов в анализе точной информации о местоположении строителей. По мере того, как технологии компьютерного зрения продолжают развиваться, точность и стабильность методов отслеживания будут постепенно улучшаться, что приведет к расширению возможностей отслеживания отдельных людей.

Распознавание. Например, Liu et al. [125] использовали силуэтный подход, в то время как Yang et al. [108] использовали SVM-подход для классификации особенностей изображения. Однако с технологическим прогрессом Янг [126] усовершенствовал метод, включив в него анализ сцен на основе данных, сохранив при этом плотные траектории. За последние несколько лет развитие сверточных нейронных сетей привело к снижению затрат на обучение алгоритмов обнаружения объектов, а также к повышению скорости и точности обнаружения. Эти усовершенствования облегчили применение алгоритмов обнаружения объектов при распознавании строительных работ. Luo et al. [43] использовали Faster R-CNN для идентификации строителей и различных объектов, изображенных на изображениях, сделанных на строительных площадках. Они использовали сеть релевантности для распознавания нескольких строительных работ, анализируя пространственные отношения между ними. Аналогичные исследования включают Fang et al. [38], которые использовали SORT и распознавание лиц для обнаружения несертифицированных работ, и Li et al. [75], которые использовали CenterNet для обнаружения строительных рабочих и объектов для оценки производительности строительства при распознавании операций по сборке арматуры. Обширные исследования в

области распознавания строительной деятельности были проведены исследовательской группой под руководством Ло Сяочуня и Ли Хэна, углубившись в объединение алгоритмов обнаружения объектов с взаимодополняющими методологиями. Например, они объединили YOLO v3 с SORT и 3D CNN для достижения пространственной локализации и распознавания строительных работ строителями [41]. В другом исследовании ученые использовали алгоритмы YOLO v3, SORT, KCF (Kernelized Correlation Filter), C3D и CRF (Conditional Random Field) в несколько этапов, что позволило распознавать строительные работы, выполняемые рабочими в группах [50]. Рассматривая пространственно-временные свойства строительной деятельности, Li et al. [76] непосредственно применили алгоритм Faster R-CNN для распознавания строительных работ на изображениях, одновременно обнаруживая строителей и определяя деятельность

# What Is Computer Vision: Benefits, Types, Libraries and more

# https://www.rishabhsoft.com/blog/computer-vision

## Что такое компьютерное зрение?

Компьютерное зрение - это область искусственного интеллекта, которая фокусируется на предоставлении компьютерам возможности интерпретировать, обрабатывать, анализировать и понимать визуальные данные из окружающего мира, из изображений и видео. Цель компьютерного зрения - воспроизвести и усилить способность человека видеть и понимать мир визуально, а также предоставить машинам способность воспринимать, рассуждать и действовать на основе визуальной информации. Оно включает методы и алгоритмы из информатики, математики, статистики и машинного обучения для извлечения значимой информации из визуальных данных и позволяет машинам распознавать объекты, обнаруживать закономерности, классифицировать изображения, отслеживать движение и выполнять другие задачи, требующие визуальной интерпретации. Компьютерное зрение имеет широкий спектр применений, от редактирования изображений и видео до

медицинской визуализации, робототехники, автономных транспортных средств, безопасности и наблюдения и многих других.

## Как работает компьютерное зрение и его ключевые компоненты?

Компьютерное зрение - это междисциплинарная область, объединяющая компьютерные науки, искусственный интеллект, математику и нейробиологические методы, позволяющие машинам интерпретировать и анализировать визуальные данные. Системы компьютерного зрения анализируют цифровые изображения и видео на высоком уровне, используя сложные алгоритмы и модели глубокого обучения.

Вот подробный обзор того, как компьютерное зрение работает с его ключевыми компонентами:

- **Получение изображения**: Первым шагом в компьютерном зрении является получение изображения или видеопотока. Это можно сделать с помощью камеры или другого устройства обработки изображений.
- **Предварительная обработка**: После получения изображения его необходимо предварительно обработать, чтобы компьютеру было легче анализировать. Это может включать уменьшение шума, улучшение изображения или цветокоррекцию.
- **Извлечение признаков**: На этом этапе компьютер анализирует изображение, чтобы идентифицировать и извлечь конкретные признаки, имеющие отношение к задаче. Это может включать обнаружение краев, углов или других форм или идентификацию объектов на изображении.
- **Распознавание объектов:** Компьютер может идентифицировать объекты на изображении после выделения соответствующих признаков. Это может включать сравнение характеристик изображения с базой данных известных

объектов или использование алгоритмов машинного обучения для распознавания узоров и форм.
- **Анализ изображения**: После идентификации объектов на изображении компьютер может проанализировать изображение более детально. Это может включать отслеживание движения предметов с течением времени, распознавание закономерностей на изображении или обнаружение аномалий или выбросов.
- **Принятие решений:** Наконец, на основе анализа изображения компьютер может принимать решения или предпринимать действия. Например, система компьютерного зрения может управлять роботизированной рукой производственного предприятия или обнаруживать потенциальные опасности в видеопотоке наблюдения.

Это упрощенный обзор того, как работает компьютерное зрение, и множество различных подходов и методик могут быть использованы для решения конкретных задач в данной области.

## Типы компьютерного зрения

Вот некоторые из наиболее распространенных типов компьютерного зрения:

- **Классификация изображений:** Это включает в себя разделение изображений на заранее определенные классы или категории, такие как животные, транспортные средства или здания.
- **Обнаружение объектов:** идентификация и локализация объектов в изображении или видеопотоке.
- **Отслеживание объектов:** мониторинг перемещения объектов в изображении или видеопотоке с течением времени.
- **Оценка позы:** Определение положения и ориентации объекта в трехмерном пространстве.
- **Семантическая сегментация:** Это включает в себя разделение изображения на сегменты и присвоение каждому компоненту метки

на основе его содержимого, такого как небо, земля, деревья или люди.
- **Распознавание действий:** идентификация и классификация действий человека в изображении или видеопотоке.
- **Восстановление изображения:** устранение шума, размытости или других искажений изображения для восстановления его первоначального качества.
- **Распознавание лиц:** распознавание человеческих лиц для разблокировки смартфонов и применения фильтров или в целях наблюдения
- **Распознавание образов:** определение формы, размера, цвета и других визуальных элементов на изображениях
- **Сегментация экземпляров:** Это похоже на семантическую сегментацию. Вместо присвоения отдельной метки каждому сегменту, каждому экземпляру объекта на изображении присваивается уникальное название.
- **Анализ движения:** Это включает в себя анализ движения объектов в последовательности видео или изображений для отслеживания их траектории, скорости и ускорения.
- **Реконструкция сцены:** предполагает создание 3D-модели сцены или объекта из нескольких 2D-изображений или видеокадров.

# Преимущества компьютерного зрения

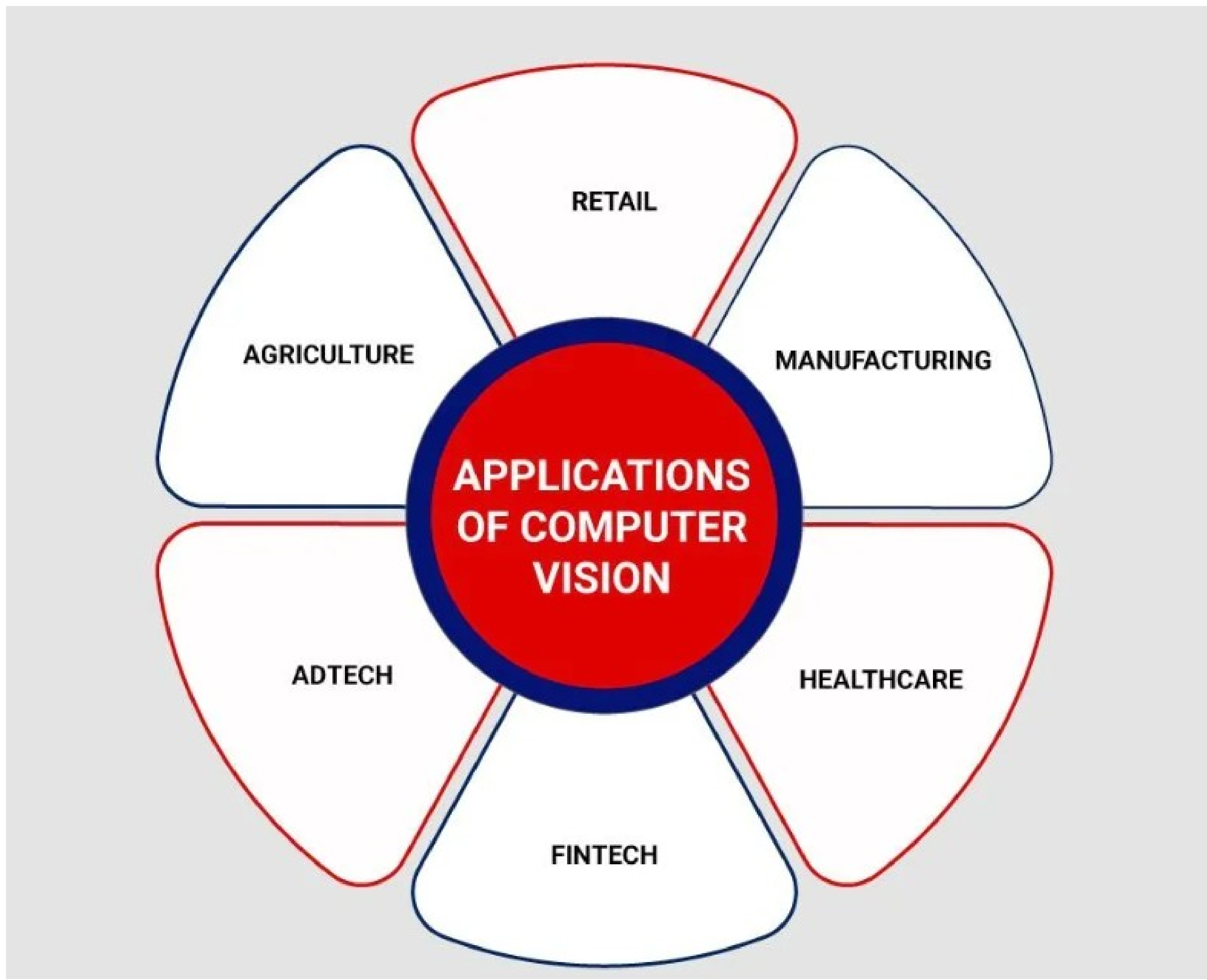
Преимущества компьютерного зрения многочисленны и далеко идущие, они находят применение в широком спектре отраслей и сфер. Вот некоторые из основных преимуществ компьютерного зрения:

- **Повышенная эффективность:** Компьютерное зрение позволяет автоматизировать повторяющиеся задачи и повысить точность, повышая эффективность в различных отраслях промышленности.
- **Автоматизация:** Компьютерное зрение позволяет автоматизировать широкий спектр задач, которые ранее требовали вмешательства

человека, таких как контроль качества на производстве, управление запасами и наблюдение за безопасностью.

- **Повышенная точность:** Алгоритмы компьютерного зрения могут анализировать визуальные данные с большой точностью, часто превосходящей человеческие возможности. Это может привести к более точным и надежным результатам, улучшению качества продукции, сокращению отходов и повышению удовлетворенности клиентов.
- **Экономия средств:** Автоматизация и повышение эффективности могут привести к значительной экономии трудовых и эксплуатационных затрат.
- **Повышенная безопасность:** Компьютерное зрение можно использовать в целях наблюдения и обеспечения безопасности для выявления угроз и предотвращения несчастных случаев.
- **Улучшение качества обслуживания клиентов:** Компьютерное зрение может помочь компаниям персонализировать свои продукты и услуги на основе предпочтений и поведения клиентов. Это может улучшить качество обслуживания клиентов и повысить их лояльность.
- **Более эффективное принятие решений:** Компьютерное зрение может предоставить компаниям информацию в режиме реального времени и аналитику данных, которые помогут им принимать более эффективные решения. Например, розничные продавцы могут использовать компьютерное зрение для отслеживания поведения покупателей и соответствующей корректировки своих маркетинговых стратегий.

# Примеры использования компьютерного зрения




Благодаря своей скорости, объективности, непрерывности, точности и масштабируемости компьютерное зрение может быстро превзойти возможности человека. Новейшие модели глубокого обучения обеспечивают точность и производительность выше человеческого уровня в реальных задачах распознавания изображений, таких как распознавание лиц, обнаружение объектов и классификация изображений. Приложения компьютерного зрения используются в различных отраслях, начиная от безопасности и медицинской визуализации и заканчивая производством, автомобилестроением, сельским хозяйством, строительством, "умным городом", транспортом и многими другими. Распространенные варианты использования компьютерного зрения включают:

## Розничная торговля

- **Отслеживание поведения:** Для отслеживания поведения покупателей в магазине, помогая определить зоны с интенсивным движением, часы пик и области, представляющие интерес. Эти

данные можно использовать для оптимизации размещения товаров и планировки магазина.

- **Механизм защиты от кражи:** обнаружение кражи и несанкционированного доступа в зоны ограниченного доступа. Его также можно использовать для выявления подозрительного поведения, такого как праздношатание или магазинная кража.
- **Управление запасами:** Автоматизация задач управления запасами, таких как отслеживание уровня запасов и местоположения продуктов. Его также можно использовать для выявления расхождений между физическими и цифровыми запасами.
- **Самостоятельные проверки:** Автоматизация процесса оформления заказа путем идентификации и отслеживания приобретаемых товаров.
- **Тепловые карты посетителей**: Обнаружение тепловых карт посетителей может помочь розничным продавцам определить наиболее популярные зоны магазина.
- **Виртуальная примерка:** Возможность виртуальной примерки таких товаров, как одежда или косметика, может улучшить качество обслуживания клиентов.

## Розничная торговля

- **Отслеживание поведения:** Для отслеживания поведения покупателей в магазине, помогая определить зоны с интенсивным движением, часы пик и области, представляющие интерес. Эти данные можно использовать для оптимизации размещения товаров и планировки магазина.
- **Механизм защиты от кражи:** обнаружение кражи и несанкционированного доступа в зоны ограниченного доступа. Его также можно использовать для выявления подозрительного поведения, такого как праздношатание или магазинная кража.
- **Управление запасами:** Автоматизация задач управления запасами, таких как отслеживание уровня запасов и местоположения продуктов. Его также можно использовать для выявления расхождений между физическими и цифровыми запасами.
- **Самостоятельные проверки:** Автоматизация процесса оформления заказа путем идентификации и отслеживания приобретаемых товаров.

- **Тепловые карты посетителей**: Обнаружение тепловых карт посетителей может помочь розничным продавцам определить наиболее популярные зоны магазина.
- **Виртуальная примерка:** Возможность виртуальной примерки таких товаров, как одежда или косметика, может улучшить качество обслуживания клиентов.

## Производство

- **Профилактическое обслуживание:** анализ данных с датчиков и камер для прогнозирования того, когда машинам потребуется техническое обслуживание. Это помогает производителям планировать техническое обслуживание в наиболее удобное время, сокращать время простоя и избегать дорогостоящего ремонта.
- **Устранение дефектов:** обнаружение дефектов и неровностей в продуктах до того, как они попадут к потребителю. Это может помочь производителям снизить затраты, связанные с отзывами продукции и неисправными компонентами, одновременно улучшая качество продукции и сокращая количество отходов.
- **Использование оборудования для обеспечения безопасности:** Мониторинг использования оборудования для обеспечения безопасности во взрывоопасных зонах, таких как строительные площадки, для обеспечения соответствия требованиям и предотвращения несчастных случаев.
- **Управление сборочной линией:** Мониторинг перемещения изделий и обеспечение их правильной сборки в процессе производства
- **Отслеживание качества продукции:** Мониторинг качества продукции во время доставки путем анализа изображений продукции по мере ее загрузки в грузовики или доставки покупателям.

## Здравоохранение

- **Анализ медицинских изображений:** Анализ медицинских изображений, таких как рентгеновские снимки, компьютерная томография и МРТ, помогает врачам в диагностике и планировании лечения. Это помогает врачам выявлять заболевания на ранней стадии, улучшать результаты лечения пациентов и снижать расходы на здравоохранение.
- **Уход за пациентами:** Дистанционное наблюдение за пациентами, особенно при обнаружении падений или других критических

инцидентов. Это повышает безопасность пациентов и позволяет медицинским работникам быстро реагировать на чрезвычайные ситуации.

- **Лабораторные тесты:** Автоматизация лабораторных тестов, которая может повысить точность и ускорить результаты.
- **Использование СИЗ:** Мониторинг использования средств индивидуальной защиты (СИЗ) в больницах и клиниках для обеспечения соответствия требованиям и предотвращения распространения инфекций.

## Финансовые технологии

- **Документирование претензий:** Анализ изображений или видеозаписей аварии или ущерба помогает страховщикам быстро и точно оценить ущерб и более эффективно обрабатывать претензии.
- **Предотвращение аварий:** Обнаружение потенциальных опасностей на дороге, таких как дорожные препятствия, пешеходы или другие транспортные средства, и оповещение водителей или принятие обходных мер, чтобы избежать аварий.
- **Обнаружение мошенничества:** Идентификация поддельных счетов-фактур путем анализа их визуальных элементов и сравнения с подлинными. Это может помочь предотвратить финансовые потери из-за мошеннических действий.
- **Биометрическое распознавание:** распознавание лиц или сканирование радужной оболочки помогает улучшить процессы безопасности и аутентификации.
- **Обработка чеков:** Считывание и обработка информации о чеках может помочь сократить время обработки и количество ошибок.

## Adtech

- **Таргетированная реклама:** анализируя изображения и видео, пользователи размещают их в социальных сетях, чтобы понять их интересы, демографические данные и поведение. Эта информация может быть использована для показа персонализированной рекламы, которая с большей вероятностью привлечет пользователей и увеличит конверсии.
- **Показы показанной рекламы:** Мониторинг показа рекламы по различным каналам, таким как онлайн-платформы, рекламные щиты

и телевидение, помогает оптимизировать расходы на рекламу и повысить эффективность кампании
- **Анализ настроений:** Анализ эмоциональных реакций пользователей на их рекламу. Это можно сделать путем анализа выражения лица, языка тела и других сигналов, чтобы определить, счастливы пользователи, грустны, сердиты или нейтральны. Эту информацию можно использовать для повышения эффективности рекламы.
- **Создание креативного контента:** Рекламодатели могут использовать генеративные модели искусственного интеллекта для создания рекламного контента, адаптированного к интересам и предпочтениям их целевой аудитории. Эти модели могут генерировать изображения, видео и другие типы мультимедиа, которые очень привлекательны и имеют больше шансов найти отклик у пользователей.

### Сельское хозяйство

- **Производство зерна:** Оптимизация урожайности путем анализа состояния посевов, выявления стрессов на посевах и прогнозирования урожайности. Это помогает фермерам совершенствовать методы управления растениеводством, сокращать отходы и повышать прибыльность.
- **Борьба с сорняками:** Выявление сорняков и нацеливание на них позволяет фермерам избирательно применять гербициды и сокращать использование химикатов, что лучше для окружающей среды.
- **Качество почвы:** Анализ качества почвы путем измерения влажности почвы, уровня питательных веществ и других факторов. Это помогает фермерам оптимизировать рост урожая, сократить потребление воды и повысить устойчивость.

## Библиотеки компьютерного зрения

Различные инструменты и платформы компьютерного зрения часто выбираются на основе конкретных потребностей проекта, а также предпочтений и опыта команды разработчиков. Вот некоторые из наиболее популярных из них.:

- **OpenCV:** OpenCV (компьютерное зрение с открытым исходным кодом) - это библиотека программных функций с открытым исходным кодом, в основном предназначенная для компьютерного зрения в реальном времени. Оно предоставляет широкий спектр алгоритмов для обработки изображений и видео, обнаружения объектов, распознавания объектов, машинного обучения и многого другого.
- **PyTorch:** PyTorch особенно хорошо подходит для задач глубокого обучения и включает в себя множество предварительно подготовленных моделей и инструментов для обучения пользовательских моделей.
- **Keras:** Keras - это высокоуровневая библиотека нейронных сетей, которую можно использовать с TensorFlow или Theano в качестве серверной части. Оно особенно хорошо подходит для быстрого создания прототипов и экспериментов и включает в себя множество предварительно подготовленных моделей для классификации изображений и других задач.
- **Detectorn2:** Эта библиотека компьютерного зрения разработана для упрощения создания приложений обнаружения объектов и сегментации. Оно включает внутреннюю поддержку для реализации алгоритмов глубокого обучения, таких как RetinaNet, Faster R-CNN, DensePose и Mask R-CNN, а также более современных алгоритмов, таких как TensorMask, Panoptic FPN и Cascade R-CNN.
- **Theano:** Theano - популярная библиотека числовых вычислений, которую можно использовать для машинного обучения и задач компьютерного зрения. Оно хорошо подходит для задач глубокого обучения и включает в себя множество предварительно подготовленных моделей и инструментов для обучения пользовательских моделей.
- **Mahotas:** Mahotas - это библиотека Python, которая предоставляет ряд алгоритмов обработки изображений для таких задач, как извлечение объектов, сегментация и фильтрация.

# Предпочтительные фреймворки, используемые Computer Vision

- **TensorFlow:** TensorFlow - это платформа машинного обучения с открытым исходным кодом, широко используемая в исследованиях и разработках в области компьютерного зрения. Оно включает в себя множество предварительно подготовленных моделей и инструментов для классификации изображений, обнаружения объектов и других задач.
- **Caffe:** Caffe - это платформа глубокого обучения, оптимизированная для задач обработки изображений. Ее можно использовать с различными языками программирования для классификации изображений, обнаружения объектов и других задач.
- **Torch:** Torch - это платформа научных вычислений, которая включает в себя множество инструментов и библиотек для обучения пользовательских моделей и особенно хорошо подходит для задач глубокого обучения.
- **MXNet:** Эта платформа глубокого обучения с открытым исходным кодом очень предпочтительна для распределенного обучения и включает множество предварительно обученных моделей для классификации изображений, обнаружения объектов и других задач.

# **Приложения компьютерного вопроса**

# **https://www.geeksforgeeks.org/applications-of-computer-vision/**

[redacted]

Компьютерное зрениеили машинное зрение — это область науки, которая позволяет компьютерам или устройствам распознавать различные объекты так же, как люди.Компьютеры необходимо научить находить объекты, а

также некоторые правила, точно так же, как вы учите ребенка распознавать объекты, но компьютеры более эффективны, поскольку обучение занимает очень мало времени.Компьютерное зрение имеет**приложения**во всех отраслях промышленности, и они заключаются в следующем:

1. **Нефть и природный газ:**Нефтегазовые компании добывают миллионы баррелей нефти и миллиарды кубических футов газа каждый день, но для того, чтобы это произошло, сначала геологи должны найти подходящее место, откуда можно добыть нефть и газ.Чтобы найти эти местоположения, им приходится анализировать тысячи различных мест, используя изображения, сделанные на месте.Предположим, если геологам придется анализировать каждое изображение вручную, сколько времени потребуется, чтобы найти наилучшую оценку?Возможно, месяцы или даже год, но из-за развития компьютерного анализа период анализа может сократиться до нескольких дней или даже нескольких часов.Вам просто нужно ввести созданные изображения в предварительно обученную модель, и она выполнит свою работу.
2. **Процесс найма:** В мире управления персоналом компьютерное зрение меняет способы приема кандидатов на работу в процессе собеседования. Используя компьютерное зрение, машинное обучение и научные данные, они могут количественно оценить «мягкие навыки» и провести раннюю оценку кандидатов, чтобы помочь крупным компаниям отобрать кандидатов.
3. **Видеонаблюдение:** Концепция пометки видео используется для пометки видео ключевых слов на основе объектов, которые обрабатываются на каждой сцене. Теперь представьте себя на месте охранной компании, которая просит обыскать подозреваемого в синем фургоне среди многочасовой видеозаписи. Вам просто нужно будет передать видео алгоритм. компьютер благодаря зрению и распознаванию объектов, поиск по видео ушел в прошлое.
4. **<strong>Строительство: </strong>Возьмем, например, электрические башни или здания, которые требуют соблюдения определенных технических требований для проверки степени ржавления и других структурных дефектов. Конечно, вручную взбираться на башню, чтобы осмотреть каждый дюйм и уголок, было бы**

**крайне чрезмерно, дорого и опасно. Полет дрона с проводами вокруг электрической башни также не кажется особенно безопасным. Итак, как вы могли бы применить здесь компьютерное зрение? Представьте себе, что если бы человек на земле сделал высокие снимки разрешения под разными углами. Затем специалист по зрению мог бы создать компьютерный пользовательский классификатор и использовать его для обнаружения дефектов и количества присутствующих ржавчины или трещин.**

5. **Здравоохранение:**За последние несколько лет индустрия здравоохранения внедрила множество технологий следующего поколения, которые включают в себя искусственный интеллект и машинное обучение.Одним из них является компьютерное зрение, которое помогает определять или диагностировать заболевания у людей или любых живых существ.
6. **Сельское хозяйство:**На сельскохозяйственных фермах начали использовать технологии компьютерного зрения в различных формах, таких как интеллектуальные тракторы, интеллектуальное сельскохозяйственное оборудование и дроны, которые позволяют эффективно и легко контролировать поля и ухаживать за ними.Это также помогает повысить урожайность и качество сельскохозяйственных культур.
7. **Военные:**Для современной армии компьютерное зрение является важным фактором, помогающим обнаруживать силы противника, а также расширяющим возможности наведения систем управляемых ракет.Он использует изображения датчиков для получения разведывательных данных о поле боя, а также для принятия тактических решений.Еще одним важным применением является применение автономных компьютерных средств, таких как беспилотные летательные аппараты и полуавтоматические автомобили с дистанционным управлением, благодаря которым необходимо перемещаться по территории страны.
8. **Промышленность:**На производстве или сборочной линии компьютерный окуляр используется для выездных проверок, выявления дефектных изделий на производственных линиях и для удаленных проверок оборудования.Технология также используется для повышения эффективности производственных линий.

9. **Автомобильная промышленность:**Это один из лучших примеров технологии компьютерного зрения, ставшей реальностью для людей.Беспилотный искусственный интеллект анализирует данные с камеры, установленной на автомобиле, для автоматизации определения полос движения, обнаружения движения и определения дорожных знаков и светофоров.
10. **Автоматическое чтение по губам:**Это одна из практических реализаций компьютерного зрения, помогающая людям с ограниченными возможностями или не умеющим говорить, она считывает движение губ и сравнивает его с уже известными движениями, которые были объяснены и использовались для создания моделей.

# Обзор современных инструментов компьютерного зрения

# https://svitla.com/blog/overview-of-modern-computer-vision-tools

## Что такое компьютерное зрение

Компьютерное зрение - это теория и технология создания машин, которые могут обнаруживать, отслеживать и классифицировать объекты.

Как научная дисциплина компьютерное зрение относится к теории и технологии создания искусственных систем,

которые получают информацию в виде изображений. Визуальные данные могут быть представлены во многих формах, таких как видеоряды, изображения с разных камер или 3D-данные медицинского сканера.

Как технологическая дисциплина, computer vision стремится применять теории и модели компьютерного зрения для создания систем компьютерного зрения. Примерами таких систем являются:

- системы управления технологическими процессами (промышленные процессы, автономные транспортные средства)
- системы видеонаблюдения
- системы организации информации (например, для индексации баз данных изображений)
- системы моделирования объектов или окружающей среды (анализ медицинских изображений, топографическое моделирование)
- системы взаимодействия (например, устройства ввода для систем взаимодействия человека и машины)

# Что такое цифровая обработка изображений

Цифровая обработка изображений - это использование компьютерных алгоритмов для обработки цифровых изображений. Как область цифровой обработки сигналов, цифровая обработка изображений имеет много преимуществ перед аналоговой обработкой. Это

позволяет применять гораздо более широкий спектр алгоритмов к входным данным и избегать таких проблем, как дополнительный шум и искажения во время обработки. Поскольку изображения определяются как двумерные (или выше), цифровую обработку изображений можно моделировать как многомерные системы. Первая цифровая обработка изображений началась в 1960 году в таких областях, как спутниковые изображения и передача изображений по проводам.

## Современные инструменты компьютерного зрения

Современное программное обеспечение для компьютерного зрения включает библиотеки для программирования, инструменты для работы с нейронными сетями и облачные решения для выполнения задач с изображениями.

Давайте рассмотрим наиболее популярные инструменты, которые сейчас широко используются на практике.

## OpenCV

OpenCV (Библиотека компьютерного зрения с открытым исходным кодом) - это библиотека функций и алгоритмов для компьютерного зрения, обработки изображений и числовых алгоритмов общего назначения с открытым исходным кодом. Библиотека предоставляет инструменты для обработки и анализа содержимого изображений, включая распознавание объектов на цифровых фотографиях (таких как лица и фигуры людей, текст и т.д.), отслеживание движения объектов, преобразование

изображений, применение методов машинного обучения и идентификацию общих элементов на различных изображениях.

Первоначально эта библиотека была разработана в Центре разработки программного обеспечения Intel. OpenCV написан на языке высокого уровня (C/C++) и содержит алгоритмы интерпретации изображений, калибровки камеры в соответствии со стандартом, устранения оптических искажений, определения сходства, анализа движения объекта, определения формы объекта и отслеживания объекта, 3D-реконструкции, сегментации объекта, распознавания жестов и т.д. Библиотека состоит из следующих компонентов:

cxcore - ядро (содержит базовые структуры данных и алгоритмы):
- базовые операции с многомерными числовыми массивами
- матричная алгебра, математические функции, генераторы случайных чисел
- Запись / восстановление структур данных в / из XML
- базовые функции 2D графики

CV - модуль обработки изображений и компьютерного зрения
- основные операции с изображениями (фильтрация, геометрические преобразования, преобразование цветового пространства и т.д.)
- анализ изображений (выделение отличительных признаков, морфология, поиск контуров, гистограммы)
- анализ движения, отслеживание объектов
- обнаружение объектов, в частности лиц
- калибровка камеры, элементы восстановления пространственной структуры

Highgui - модуль для ввода/вывода изображений и видео, создания пользовательского интерфейса
- захват видео с камер и из видеофайлов, чтение/запись статических изображений.
- функции для организации простого пользовательского интерфейса (все демонстрационные приложения используют HighGUI)

Cvaux - экспериментальные и устаревшие функции
- пробелы. Зрение: стереокалибровка, самокалибровка
- поиск стереосопоставления
- нахождение и описание черт лица

## Tensorflow

На сегодняшний день это самая популярная библиотека машинного обучения и глубокого обучения. ее популярность быстро возросла и превзошла существующие библиотеки благодаря простоте API. Google выпустила ее в ноябре 2015 года.

Написано на Python, но теперь есть и порт JavaScript (tensorflow.js). Это дополнение связано с растущей популярностью JavaScript после выпуска Node.js.

TensorFlow - это бесплатная библиотека с открытым исходным кодом для потоков данных и дифференциального программирования. Это библиотека символьной математики, которая также используется для приложений машинного обучения, таких как нейронные сети.

В TensorFlow 2.0 по умолчанию в качестве режима выполнения модели используется оперативное выполнение. То есть вычисление конкретных значений происходит попутно для построения полного

вычислительного графика. Это упрощает отладку модели и устраняет необходимость в шаблонном коде.

Вы можете использовать стандартные структуры Python в качестве структур данных. Вы можете быстро проверять гипотезы и легко отлаживать код на небольших моделях и небольших объемах данных. Eager execution также поддерживает ускорение GPU и распределенные вычисления на многих машинах.

TensorFlow 2.0 облегчает реализацию предварительно обученных моделей, которые настроены на распознавание изображений и речи, обнаружение объектов, рекомендации, усиленное обучение и т.д. Такие эталонные модели позволяют вам использовать лучшие практики "из коробки" и служат отправной точкой для разработки ваших собственных высокопроизводительных решений.

## CUDA

CUDA (Compute Unified Device Architecture) - это аппаратно-программная архитектура для параллельных вычислений, которая может значительно повысить производительность вычислений за счет использования графических процессоров Nvidia.

CUDA SDK предоставляет возможность включать процедуры вызова C, работающие на графических процессорах Nvidia, в текст C. Это делается с помощью команд, написанных на специальном диалекте C. Архитектура CUDA дает разработчику возможность по своему усмотрению получать доступ к набору инструкций графического ускорителя и манипулировать им.

Оригинальная версия CUDA SDK была представлена 15 февраля 2007 года. CUDA API основан на языке C.

Благодаря ускорению CUDA приложения могут добиться повышения частоты кадров интерактивного видео.

- Сегментация по уровням с помощью CUDA
- Сегментация видео с помощью CUDA
- Мультиклассовая реализация SVM в CUDA
- Обнаружение пешеходов
- Flowlib: плотный оптический поток
- Байесовский оптический поток
- Машинное обучение и обработка данных
- Аппаратное обеспечение эффективного распространения убеждений
- Быстрый поиск k-ближайших соседей с помощью графического процессора

С помощью CUDA задачи компьютерного зрения могут решаться быстрее, а технологии выходят на новый уровень; например, возможность автоматического управления транспортными средствами, обработки больших объемов фото- и видеоинформации, медицинских видеоданных и так далее.

## Theano

Theano - это быстрая цифровая библиотека Python, которая может работать на CPU или GPU. Она была разработана командой LISA (ныне MILA) из Монреальского университета в Канаде. Theano - оптимизирующий компилятор для обработки и оценки математических выражений, особенно матричных.

Вот ключевые методы, операции и структуры данных, поддерживаемые Theano:

- тензоры через структуру numpy.ndarray и поддержка многих тензорных операций
- разреженные матрицы с помощью scipy. {csc, csr, bsr} _матричные структуры и поддержка ряда операций с ними
- возможность создавать новые операции с графиками в рабочем режиме
- многочисленные операции преобразования графиков
- Поддержка языка Python для версий 2 и 3
- Поддержка графических процессоров (CUDA и OpenCL)
- Стандартная поддержка базовых подпрограмм линейной алгебры (BLAS) для процедур линейной алгебры

## Keras

Keras - это библиотека Python для глубокого обучения, которая сочетает в себе функции других библиотек, таких как Tensorflow, Theano и CNTK. Keras имеет преимущество перед конкурентами, такими как Scikit-learn и PyTorch, поскольку работает поверх Tensorflow.

Keras может работать поверх TensorFlow, Microsoft Cognitive Toolkit, Theano или PlaidML. Разработанный для быстрого экспериментирования с глубокими нейронными сетями, он ориентирован на удобство использования, модульность и расширяемость.

Keras - это API, предназначенный для людей, а не для машин. Keras следует лучшим практикам снижения

когнитивной нагрузки: он предлагает согласованные и простые API и минимизирует количество действий пользователя, необходимых для обычных случаев использования.

## Scikit-учиться

Эта популярная библиотека машинного обучения построена на NumPy, SciPy и matplotlib. Она ориентирована на алгоритмы ML.:

- Обучение под наблюдением
- Обучение без учителя
- Линейная регрессия
- Логистическая регрессия
- Вспомогательно-векторная машина (SVM)
- Наивный байесовский классификатор
- Повышение градиента
- Кластеризация
- Метод K-means

Эта библиотека менее развита, чем Tensorflow, однако предоставляет простые и эффективные инструменты для обнаружения и анализа данных.

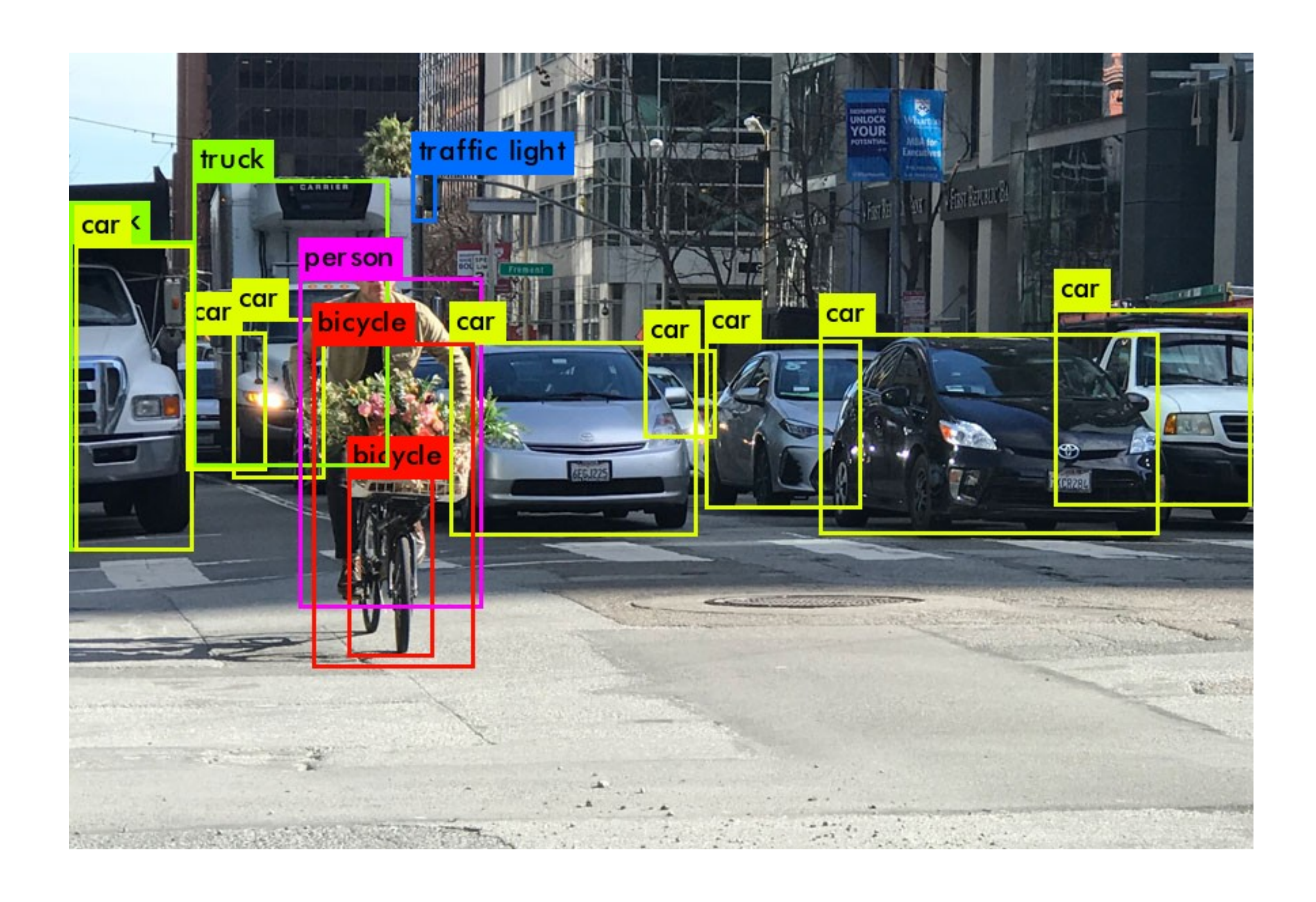
## YOLO

“Вы смотрите только один раз” или YOLO - это система обнаружения объектов, разработанная специально для обработки данных в режиме реального времени. YOLO - это передовая система обнаружения объектов в реальном времени, разработанная Джозефом Редмоном и Али Фархади из Вашингтонского университета. На их веб-сайте вы можете найти реализации SSD300, SSD500,

YOLOv2 и Tiny YOLO. Их алгоритм применяет нейронную сеть ко всему изображению, и нейронная сеть делит изображение на сетку и помечает области обнаруженными объектами.

- Очень быстрая и высокопроизводительная обработка данных в режиме реального времени.
- Прогнозы делаются из одной сети.
- YOLO является обобщенным. Он превосходит другие методы при обобщении на основе естественных изображений.
- Методы определения региона ограничивают классификатор конкретным регионом. YOLO обеспечивает доступ ко всему изображению при прогнозировании границ.
- YOLO обнаруживает по одному объекту на ячейку сетки.

Рис. 1. Обнаружение объектов YOLO.
Источник изображения




# Облачные инструменты компьютерного зрения

## API Google Cloud и Mobile Vision

Google предоставляет AutoML Vision в виде облачного сервиса. Это позволяет автоматизировать обучение вашим собственным моделям машинного обучения. Необходимо загружать изображения и обучать пользовательские модели изображений с помощью простого в использовании графического интерфейса AutoML Vision. AutoML позволяет оптимизировать ваши модели с точки зрения точности, задержки и размера; и экспортировать их в ваше приложение в облаке или на множество устройств.

API Vision от Google Cloud предлагает мощные предварительно подготовленные модели машинного обучения с помощью REST и RPC API. Он присваивает изображениям метки и быстро классифицирует их по миллионам предопределенных категорий. Google Vision API позволяет нам распознавать объекты и лица, читать печатный и рукописный текст и создавать ценные метаданные в каталоге изображений.

Vision API Google Cloud позволяет разработчикам анализировать изображения и контекстные данные с помощью самообучающихся и развивающихся моделей машинного обучения ин простой REST API. Мы можем получать контекстную информацию об изображении и классифицировать изображения по категориям и подкатегориям, достигая глубокого уровня детализации информации. Давайте рассмотрим наиболее интересные функции Vision API от Google Cloud.:

- Обнаружение тегов (определение категорий внутри изображения).
- Обнаружение явного контента (распознавание непристойного или насильственного содержания на изображении).
- Распознавание популярных логотипов.
- Распознавание географических ориентиров: (естественных и искусственных сооружений).
- Оптическое распознавание символов (обнаружение и извлечение текста внутри изображения, API распознает язык текста).
- Распознавание лиц (распознает несколько лиц на изображении, а также другие атрибуты).
- Атрибуты изображения (определение общих атрибутов изображения, таких как доминирующие цвета).

# Amazon Rekognition

Amazon Rekognition позволяет легко встраивать в приложение аналитику изображений и видео с помощью глубокого обучения. Сервис может распознавать объекты, людей, текст, сцены и действия, а также обнаруживать неприемлемый контент. Кроме того, Amazon Rekognition точно анализирует и распознает лица на изображениях и видео клиентов. Amazon Rekognition основан на точной, масштабируемой технологии глубокого обучения. Amazon Rekognition - это простой API, который быстро анализирует любые изображения и видео, хранящиеся в Amazon S3.

*Предварительно обученные алгоритмы*

- Распознавание знаменитостей на изображениях
- Определение атрибутов лица на изображениях, включая пол, возрастной диапазон, эмоции
- People's Pathing позволяет отслеживать людей с помощью видео. Рекламируемое использование этой возможности заключается в отслеживании спортивных игроков для анализа после игры.
- Обнаружение и классификация текста на изображениях
- Обнаружение небезопасного визуального контента

*Алгоритмы, которые пользователь может обучить на пользовательском наборе данных*:

- SearchFaces позволяет пользователям импортировать базу данных изображений с предварительно помеченными лицами, обучать модель машинного обучения на основе этой базы данных и предоставлять модель в виде облачного сервиса с API.
- Проверка пользователя по лицу

# Microsoft Azure Computer Vision API

Этот современный облачный API позволяет разработчикам использовать сложные алгоритмы для извлечения ценной информации из изображений для классификации и обработки визуальных данных. Служба компьютерного зрения Azure предоставляет разработчикам доступ к передовым алгоритмам, обрабатывающим изображения и возвращающим информацию. Использовать Computer Vision можно либо с помощью встроенного SDK, либо напрямую с помощью REST API:

- Визуальные возможности тегов
- Обнаружение объектов
- Классификация изображения
- Опишите изображение
- Распознавание лиц
- Определение типов изображений
- Обнаружение контента, специфичного для конкретной предметной области
- Узнайте, что представляет интерес
- Извлечение текста из изображений
- Умеренное содержание изображений
- Анализ видео в реальном времени

Одним из преимуществ этой облачной системы является ее гибкая ценовая политика, включая бесплатный тестовый доступ.

# Заключение

В заключение мы можем сказать, что теперь обработка изображений и компьютерное зрение больше не являются экзотическими или узкоспециализированными отраслями компьютерных систем. Все чаще мы будем использовать эти методы в очень широком ассортименте продуктов.

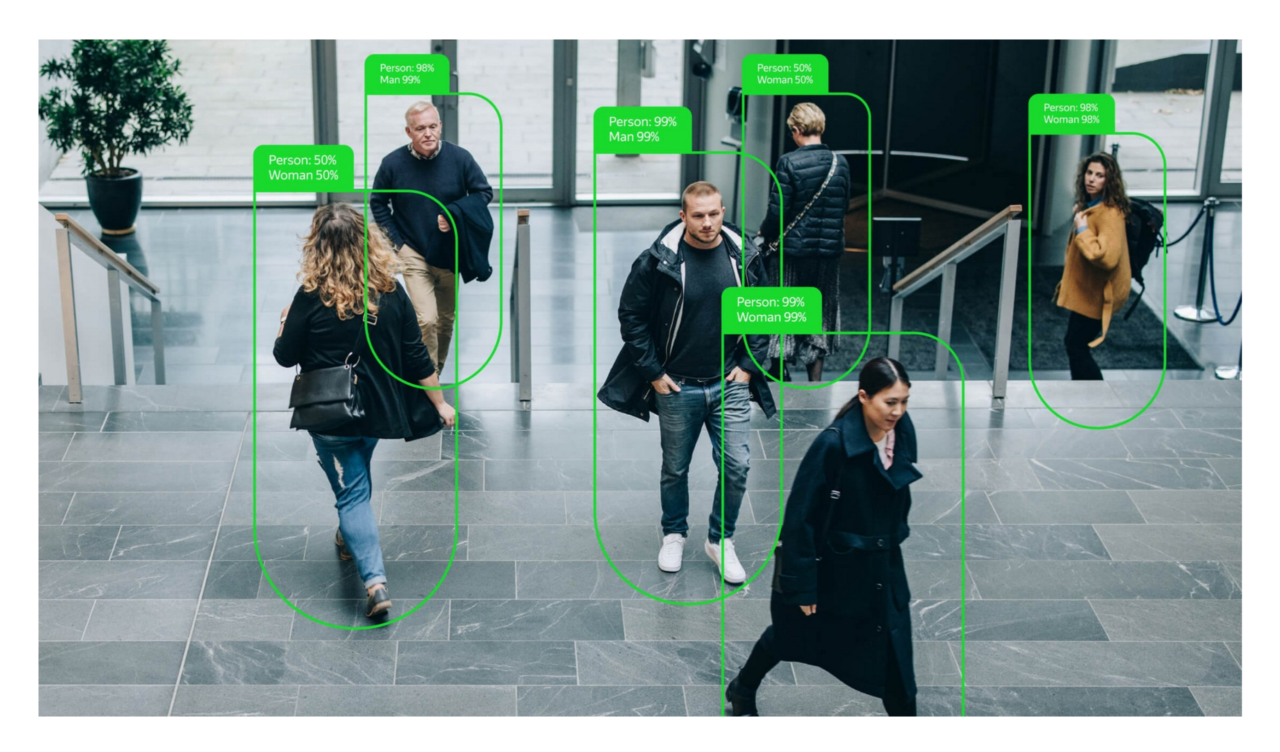
Все больше и больше разработчиков будут использовать библиотеки и облачные решения в мобильных приложениях, системах умного дома, Интернете вещей и транспортных системах. Облачные решения предоставляют вам возможность работать с хорошо подготовленными и мощными вычислительными ресурсами, в то время как клиентские фреймворки и библиотеки дают вам возможность обрабатывать эти решения непосредственно на устройстве, без необходимости доступа в Интернет.

https://cloud.yandex.ru/ru/blog/posts/2022/05/computer-vision

Компьютерное зрение (computer vision, CV) — это научная область, связанная с анализом изображений и видео. Специалисты ищут всё более совершенные способы обучить компьютер правильно видеть и извлекать информацию из увиденного. Казалось бы, это так просто: научить искусственный интеллект распознавать визуальные образы. Но не тут-то было.

Компьютер видит не так, как люди. У него нет нашего жизненного опыта и способности так же легко идентифицировать объекты на изображении. Он не способен отличить дом от дерева, не имея каких-то исходных данных. Чтобы научить компьютер видеть и понимать, что находится на изображении, люди используют технологии машинного обучения.

Для этого собирают большие базы данных, из которых формируют дата-сеты. Выделив признаки и их комбинации для идентификации похожих объектов, можно натренировать модель машинного обучения распознавать нужные типы закономерностей. Конечно, даже после загрузки нескольких дата-сетов модели могут неверно распознавать некоторые объекты. Если такое случается, модели дообучают на новых наборах данных.




ML-проекты могут потреблять значительное количество ресурсов, поэтому обучение часто проходит с использованием облачной инфраструктуры.

Например, сервис Yandex DataSphere. Благодаря ему можно быстро организовать цикл машинного обучения в облаке и запускать ML-модели в эксплуатацию без необходимости управлять инфраструктурой.

В этой статье мы расскажем:

- Компьютерное зрение на службе у человека
- Где используется компьютерное зрение
- Как обучается модель компьютерного зрения

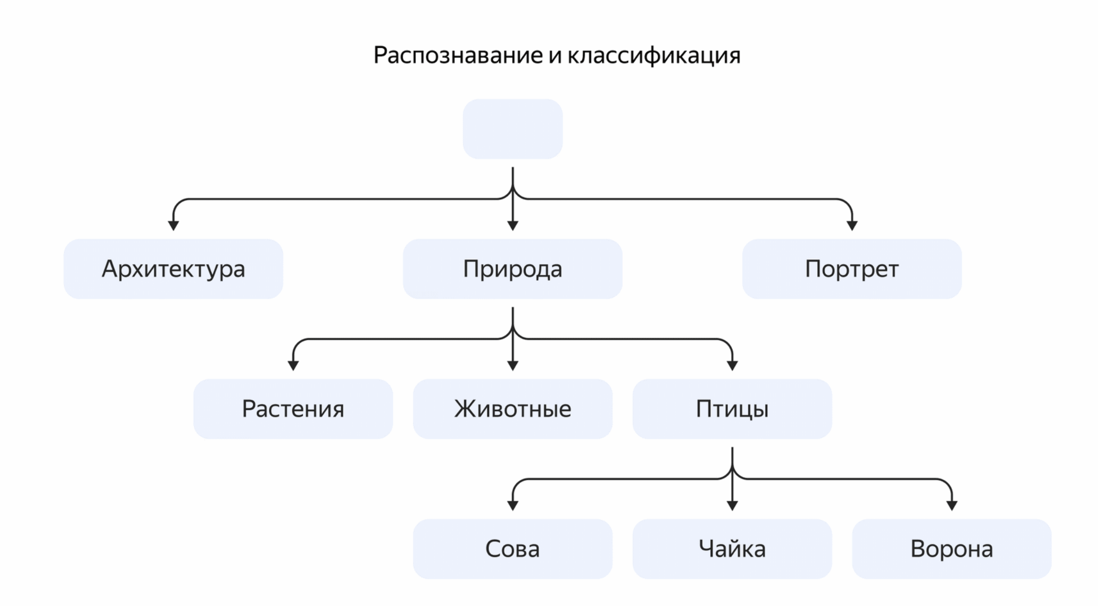
Когда именно возникла идея компьютерного зрения, точно неизвестно. Считается, что одной из первых статей в научном журнале, так или иначе затрагивающих эту тему, была публикация Receptive fields of single neurons in the cat's striate cortex. Её написали в 1959 году Дэвид Хьюбел и Торстен Визель, нейропсихологи из медицинского института Уилмера в США. В своей работе они изучали свойства нейронов зрительной коры кошек. И заметили, что зрительный опыт способен влиять на эти нейроны.

В это же время в схожем направлении думал и профессор Массачусетского технологического института Лоуренс Робертс. Он был уверен, что можно научить компьютер видеть, создал систему распознавания форм предметов и защитил в МТИ докторскую диссертацию по этому направлению.

И действительно, очень скоро компьютеры научились распознавать статичные изображения, в 80-х годах появились теории систем распознавания движущихся объектов на видео, а в 90-х учёные работали над прототипами беспилотных автомобилей.

Люди поняли, что технологии компьютерного зрения могут использоваться везде, где есть какие-либо изображения. А благодаря развитию интернета появились массивы оцифрованных изображений, которые можно было анализировать. Это позволило обучить компьютер. Сначала — распознавать текст на сканах

документов. Постепенно задачи усложнялись. Так технологии дошли до обнаружения лиц и их почти мгновенного распознавания на фото и видео.

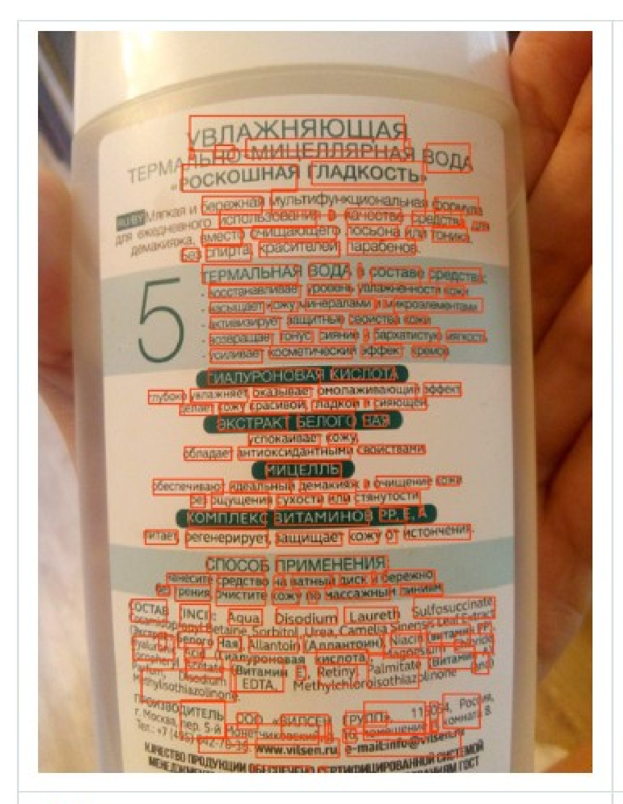
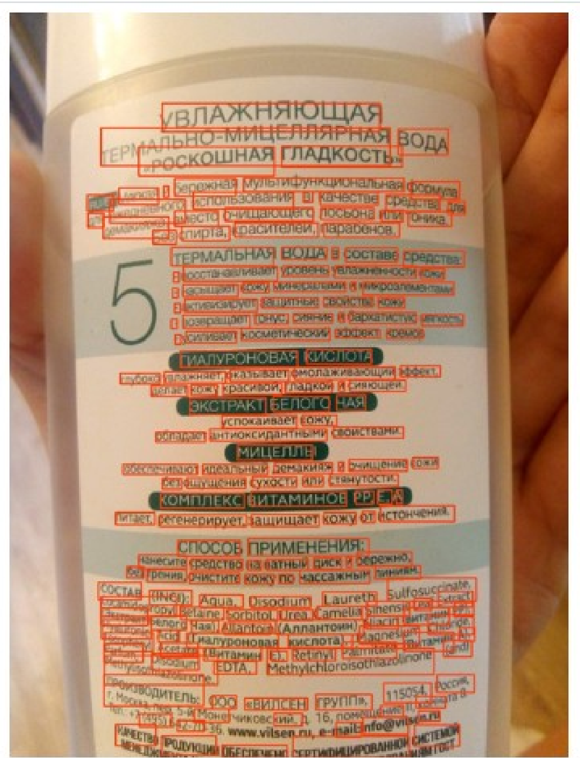



## Где используется компьютерное зрениеГде используется компьютерное зрение

### Распознавание текстаРаспознавание текста

Распознавание текста — это один из самых первых проектов Yandex Cloud. Сложность задачи заключается в том, что у изображений разное качество. Нестандартные шрифты, языки, дизайн — компьютер должен уметь читать любые слова и предложения. Если на начальном этапе программа распознавала текст отсканированного документа, то позднее добавились фотографии этих же документов (порой нечёткие), ценники, этикетки и т. д. Причём пользователям нужно, чтобы текст распознался быстро и точно. А если он ещё и с иностранного языка будет переведён на русский или любой другой — то вообще хорошо.

Решение Яндекса позволяет на лету распознавать, переводить и искать в интернете текст с самых разных объектов. Подробнее об этом можно прочесть в статье о том, как Яндекс создавал технологию оптического распознавания текста OCR.

<table>
<tr>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>2018<br>Увлажняющая<br>п НО -мичеллярная<br>роскошная гладкость вода.<br>бережная мультифункциональная<br>формула<br>использования В качестве средства<br>SI ДЛЯ<br>вместо очищающего лосьона или<br>тоника<br>Без спирта, красителей, парабенов<br>...</td>
<td>2019<br>УВЛАЖНЯЮЩАЯ<br>ТЕРМАЛЬНО-МИЦЕЛЛЯРНАЯ ВОДА<br>«РОСКОШНАЯ ГЛАДКОСТЬ»<br>AUBY Мягкая и бережная<br>мультифункциональная формула<br>для ежедневного использования в<br>качестве средства для<br>демакияжа, вместо очищающего<br>лосьона или тоника.<br>Без спирта, красителей, парабенов<br>...</td>
</tr>
</table>

## Поиск изображенийПоиск изображений

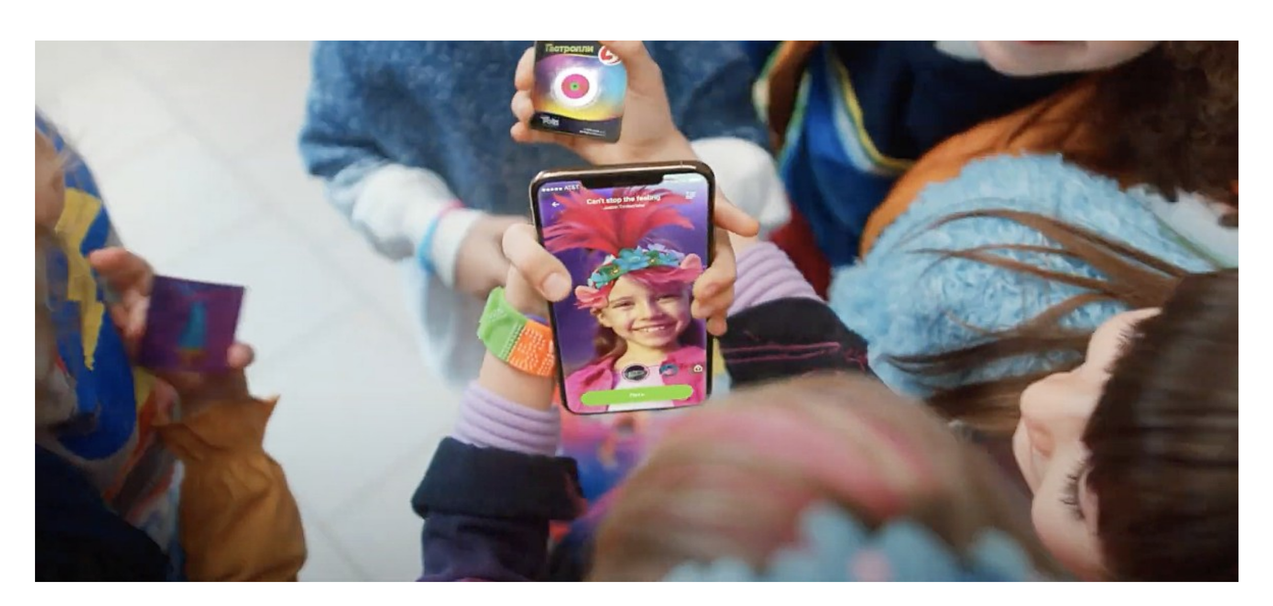
Наверняка вы не раз пользовались поиском Яндекса по картинкам. На раннем этапе развития технологии искать приходилось с помощью текстового описания, но теперь достаточно сфотографировать объект, чтобы поисковая система выдала похожие варианты. Это стало возможным благодаря технологии машинного обучения с использованием многослойных нейронных сетей, которые самообучаются на больших дата-сетах с изображениями.

Чтобы такой поиск был эффективен, индексируются изображения, ранее загруженные в интернет. Для каждого изображения компьютер строит цифровые представления, формирует структуру данных: по ней будет вестись поиск.
С изображением, которое загрузил пользователь, происходит то же самое, после чего в базе данных ищутся такие же или похожие картинки. Умная камера Яндекса умеет находить по фото даже товары в интернет-магазинах. Из других решений компьютерного зрения, предлагаемых Yandex Cloud, можно выделить определение качества изображения и кодирование файла в Base64.

## Модерирование изображенийМодерирование изображений

С каждым годом пользователи загружают всё больше контента. Контролировать миллионы фото и видео вручную невозможно. Но необходимо фильтровать их, выявляя материалы для взрослых и шок-контент. Здесь снова приходит на помощь компьютерное зрение. Решение модерирования изображений, предложенное Yandex Cloud, позволяет быстро обнаруживать публикацию нежелательного контента: материалов для взрослых, объектов, защищённых авторским правом, и т. д.

Почему это важно? В качестве примера опишем реальный кейс. Дети пользуются многими интерактивными мобильными приложениями. Например, приложение Тролли. Караоке с элементами дополненной реальности за несколько месяцев после запуска установили более 2 млн раз, туда загрузили более 15 тыс. видеозаписей. Вручную отслеживать такой поток контента практически невозможно.



На помощь пришло объектное хранилище Yandex Cloud и сервис Yandex Vision, отвечающий за распознавание элементов, которых не должно быть в кадре. В основе сервиса лежит анализ изображений с помощью моделей машинного обучения. Компьютерное зрение оценивало, насколько загруженные ролики соответствуют правилам, распознавало текстовый ненормативный контент, появляющийся на видео. А за проверку аудиодорожки отвечал SpeechKit — сервис, на основе которого работает Алиса. Получившаяся умная система премодерации защитила детей от нежелательного контента.

## Изучение космосаИзучение космоса

Снимки со спутников и телескопов долгое время анализировали люди. За многолетнюю историю наблюдения за космосом накопилось огромное количество данных из самых разных источников. И в этих данных может содержаться немало ценной информации, которую просто не заметили. Например, в декабре 2017 года астрономы, используя искусственный интеллект, проанализировали данные, собранные телескопом «Кеплер». Компьютерное зрение увидело то, что не заметил человеческий глаз: солнечную систему с восемью планетами.

## Развитие биометрииРазвитие биометрии

Изображение лица, радужная оболочка глаза, отпечаток пальца могут служить отличными идентификаторами человека. Но для этого нужна точность и производительность компьютера. Благодаря совершенствованию алгоритмов компьютерного зрения стала возможной разблокировка телефона по отпечатку

пальца, платежи FacePay. Более того, компьютерное зрение способно распознавать даже лица людей, которые носят маску.

Яндекс представил технологию распознавания лиц ещё в 2012 году, а в 2019 году запустил облачный сервис Yandex Vision с технологиями компьютерного зрения для сторонних разработчиков. С помощью сервиса можно отметить людей на фотографии, найти все фото с портретами, автомобильные номера или шаблоны документов (например, паспорта) в большом дата-сете.

### 3D-анализ3D-анализ

Реконструкция зданий и других объектов по изображению — это ещё одна интересная задача, которая решается благодаря нейронным сетям. Загрузив в компьютер фотографии разрушенного объекта, на выходе можно получить реконструированную модель. Хорошим примером может стать решение Instance Segmentation. В данном случае технологии компьютерного зрения упрощают перенос любых графических данных в 3D-модели и оптимизируют создание новых продуктов. Пока не удалось добиться, чтобы данные передавались в режиме real-time. Но в перспективе это решение может перерасти в онлайн-сканирование объектов и пространств.

### Укрепление сельского хозяйстваУкрепление сельского хозяйства

Качество и количество урожая, увеличение поголовья скота зависят от многих факторов. Быстро проанализировать их способен только компьютер. Сбор урожая тоже можно поручить искусственному интеллекту. Например, испанская компания Agrobot создала автоматический сборщик клубники. Устройство умеет самостоятельно ориентироваться в пространстве и оценивать зрелость ягод с помощью технологий компьютерного зрения. А решение Sonoma позволяет почти без участия человека выращивать огурцы. Обученный компьютер через систему камер и датчиков следит за состоянием почвы и рассады и управляет ирригацией, подкормкой, температурным режимом и другими параметрами, необходимыми для созревания урожая.

### ВидеоаналитикаВидеоаналитика

На дорогах, в метро и наземном транспорте, в офисах, подъездах и куче других мест вы наверняка видели камеры. Они отвечают за контроль движения автомобилей, мониторинг людских потоков, оповещение об инцидентах и т. д. Современные технологии компьютерного зрения позволяют идентифицировать преступника в большой толпе, вычленить нарушителя в потоке машин, увидеть брак на конвейерной ленте завода. Благодаря технологиям глубокого обучения компьютеры теперь умеют проводить геологические исследования не хуже людей!

Говоря о компьютерном зрении, можно вспомнить системы беспилотных автомобилей, контроль размеров или наполняемости упаковок на пищевом производстве, мониторинг зазоров и других характеристик деталей на автосборочных конвейерах и т. д.

После обучения компьютерные системы способны на многое. Но как проходит обучение?

## Как обучается модель компьютерного зренияКак обучается модель компьютерного зрения

Для обучения компьютера собирается массив данных, с помощью которого можно выделить характеристики, присущие тем или иным объектам. Чтобы компьютер мог находить котиков на фото, ему нужна обучающая выборка. Это дата-сет с изображениями котов (положительные примеры), который разбавлен картинками без котов (отрицательные примеры).

Во время обучения компьютер анализирует изображения, выделяя признаки и комбинации признаков, которые позволяют понять, что на картинке кот. Чем больше и разнообразнее коллекция картинок, тем точнее искусственный интеллект научится распознавать объекты. Если машинное обучение прошло успешно, компьютер справится с задачей распознавания. Если процент ошибок высок, можно провести дообучение на других дата-сетах.

## ElectroNeek: опыт применения Yandex Vision

История успеха

Обучающие дата-сеты обычно очень крупные, поэтому обучение модели компьютерного зрения может занимать много времени. Чтобы ускорить процесс, такого рода задачи разумнее выполнять в облаке с помощью виртуальных графических процессоров. Производительность GPU выше стандартных CPU, а масштабирование позволяет быстро получить объём ресурсов, необходимый для полноценного обучения модели. Например, для создания систем компьютерного зрения можно использовать сервис ML-разработки Yandex DataSphere, в котором есть инструменты и динамически масштабируемые ресурсы, необходимые для полного цикла разработки машинного обучения.

Как выглядит полный цикл машинного обучения в облаке

Если сильно упрощать, то любая система компьютерного зрения при анализе изображения проходит через три этапа:

- **Классификация изображения.** Компьютер присваивает картинке некий класс из заранее известных.
- **Локализация конкретного объекта.** Чем больше объектов на изображении, тем сложнее требуется нейросеть.
- **Построение изображения.** Программа убирает шум и повышает качество картинки, выделяя доминирующие и малозначительные объекты.

Каждый этап сложен, ведь компьютер не понимает, что на изображении важное, а что не очень. Поэтому сначала картинка проходит через внутренние алгоритмы, заложенные разработчиками. Так компьютер получает начальные сведения об изображении. Затем компьютер находит объекты и их границы. Для этого есть разные способы. Например, можно использовать размытие по Гауссу, когда изображение размывают несколько раз, выявляя самые контрастные фрагменты. Эти значимые места в дальнейшем считаются объектами.

Значимые места компьютер при помощи любого из популярных алгоритмов (например, SIFT, SURF, HOG) описывает в числовом виде. Такая запись называется дескриптором. Она позволяет полно и точно сравнить фрагменты изображений, не используя сами фрагменты. Но так как сравнение — это тяжёлая вычислительная операция, то дескрипторы кластеризуют: делят на группы. В каждой группе находятся дескрипторы разных изображений, но с общими характеристиками.

Кластеризация и следующее за ним квантование (обобщение) дескрипторов уменьшает объём данных, которые приходится обрабатывать компьютеру. А заодно позволяет быстрее распознавать объекты и сравнивать изображения.